In combination with PCR amplification, restriction enzyme analysis forms a rapid and general technique to identify gene lesions that change a restriction enzyme site. After the PCR product has been digested with a known restriction enzyme and subjected to electrophoresis, a specific pattern of bands appears on the gel. More than 100 different restriction enzyme recognition sequences are known, but because only one in six of the basepairs in the human haploid genome are covered by a naturally occurring restriction enzyme recognition site [1], >80% of mutations are not directly detectable. In such instances, appropriate restriction sites can be artificially created by siting a single nucleotide mismatch in the 3' end of an oligonucleotide primer immediately adjacent to the mutation site [2]. After incorporation of the primer into a PCR product, the (usually common) variant nucleotide of one allele matches the sequence of the primer to introduce a restriction enzyme recognition site, and the corresponding nucleotide of the other allele does not. After digestion of the PCR product, cleavage will occur completely, partially, or not at all, and this identifies which alleles were present in the template DNA. A limitation of the method is that a restriction site must be created within the primer in the confines of the genomic DNA sequence; for some mutations, there may not be a potential restriction site to engineer.

We have adapted this approach to detect rare mutations at mutation-prone CpG sites [3]. Artificial restriction sites can be established in known DNA sequences by using either sense or antisense (or both) PCR primers to force the site through mismatch of the 3'-terminal base with the target. In each case, forcing or not, both primers can be sited directly adjacent to the target CpG dinucleotide, enabling a general basis for identifying mutations at this mutation-prone site, when flanked by any combination of two bases, one 5' and the other 3'. This establishes a final common pathway for analysis of de novo mutation at any CpG site, which we term CpG-PCR.

In the present study, we have extended the CpG-PCR technique to screen pooled DNA samples for rare CpG mutations by combining a mutant enrichment strategy with forcing a TaqI restriction site in the normal sequence. Following an established principle [4], the addition of TaqI to the PCR reaction during the early cycles of amplification destroys the normal-strand templates containing the forced TaqI site of the CpG-PCR and leaves the mutant strands unrestricted for selective amplification in subsequent cycles. We have demonstrated the technique by using a CpG mutation in the LDL receptor gene, E80K [5]. E80K heterozygotes were identified in a large-scale scan of FH probands for LDL receptor mutations previously undertaken in this laboratory [6].

Normal and mutant DNA stock solution concentrations were equalized after spectrophotometric quantitation, and four dilutions of mutant: normal DNA were tested: 1:V10, 1:10, 1:100 V10, and 1:100. PCR constituents were as described previously [3], except that dimethyl sulfoxide (DBD), 10 mL/L, replaced 0.05% W1. The sense oligonucleotide used to force a TaqI recognition site in the normal sequence involved a mismatched T against target A at the 3' terminus of the primer (indicated in boldface type); sense, 5'-GGCCAGTGGACTGCGACAACGGCT-3'.

In so doing, the next three inserted bases C, G, and A in the normal sequence create a recognition site T/CGA for the restriction endonuclease TaqI in the extension product. The mutation G→A destroys the TaqI site. The antisense primer was sited downstream of the CpG site, giving a
product of 69 bp: antisense, 5’-ATAGCAAGGCGAGGACCACCTTAC-3’.

Samples were amplified in 20 µL with 20 µL of mineral oil overlay in Omniplates on a Hybaid Omnimene (Hybaid). Conditions for standard PCR amplification were as described by Hobbs et al. [7].

In the mutant enrichment PCR, 1 µL of a TaqI digestion mixture was added to each PCR reaction after the first and third amplification cycles. The first cycle was 94 °C for 5 min, 55 °C for 1 min, and 70 °C for 1 min. After the first addition, the reactants were incubated at 65 °C for 60 min in situ on the PCR block. The PCR program was then run for an additional two cycles as follows: 94 °C for 1 min, 55 °C for 1 min, and 70 °C for 1 min. An additional 1 µL of TaqI digestion mixture was then added, followed by an incubation at 65 °C for 60 min. The PCR program then ran for the remaining 32 cycles as in the previous two. The TaqI digestion mixture contained all of the PCR constituents, with the exception of the oligonucleotides, at the same concentration as in the PCR reaction so that on the addition of digestion mixture, the PCR constituent concentrations remained unchanged. Five units of TaqI were added to the 20 µL of PCR product on each 1-µL addition of digestion mixture, so that the final incubation contained 10 U of enzyme in a volume of 22 µL. After PCR, 7 µL of PCR product was digested for 2 h with 5 U of TaqI restriction endonuclease (New England Biolabs) as directed. PCR products and digests were analyzed by microplate diagonal gel electrophoresis (MADGE) as described previously [3].

E80K heterozygotes were identified by sequencing. Samples were amplified by PCR under the following conditions: one cycle at 96 °C for 5 min, 35 cycles at 96 °C for 1 min, 57 °C for 1 min, 72 °C for 1 min, and one cycle at 72 °C for 5 min. Exon 3 of the gene was amplified by using the following primers, giving a PCR product of 176 bp: sense, 5’-TGACAGTTCAATCTGTCTCTTG-3’; antisense, 5’-ATAGCAAGGCGAGGCCACCTTAC-3’.

Samples were amplified in 50-µL reactions with 20 µL overlay of mineral oil, using reagents as described above, with W1 as enhancer. PCR products were purified on Microcon-100 columns (Amicon) as described in the cycle sequencing protocol. Purified PCR products were sequenced in both directions by using the Perkin-Elmer ABI fluorescent cycle sequencing kit with AmpliTaq FS DNA polymerase and fluorescently labeled dideoxy chain terminators (Applied Biosystems). Usually 1 µL of the purified PCR product (diluted 1:10 in water) was used as template in a 10-µL sequencing reaction containing 4 µL of manufacturer’s ready reaction solution and 1.6 pmol of either of the above primers. The remaining volume was made up with water and overlaid with 20 µL of mineral oil. Cycle sequencing was carried out on a Perkin-Elmer DNA thermal cycler under the following conditions: 25 cycles at 96 °C for 30 s, 50 °C for 15 s, and 60 °C for 4 min. The products of the 10-µL cycle sequencing reaction were extracted in 50 µL of 95% (by vol) ethanol and 2 µL (3 mol/L) sodium acetate (pH 4.6) and washed in 70% (by vol) ethanol before resuspension in 5 µL of a loading buffer comprising 25 mmol/L EDTA (pH 8.0) and 50 g/L Blue dextran in deionized formamide (5:1 formamide to EDTA/Blue dextran). Samples were heated to 95 °C for 2 min and kept on ice until loading. Samples (1.5 µL each) were loaded onto a 36 cm 4.0% acrylamide gel containing 6 mol/L urea (Sequagel XR, National Diagnostics) and electrophoresed at 48 °C for 3.5 h. The sequence was read automatically by using Sequencing Analysis software version 3.0 (Applied Biosystems).

Figure 1 shows the PCR products obtained by forcing a TaqI site in the normal sequence at the LDL receptor gene CpG mutation sites E80K in exon 3. The products obtained in a standard PCR (Fig. 1A) are shown in comparison with those obtained after the addition of TaqI to the PCR reaction after the first and third cycles of amplification, i.e., mutant enrichment (Fig. 1B). There was no difference in the amount of product yield in standard and mutant-enriched PCRs, showing that although the addition of TaqI in the early stages of amplification destroyed normal template, sufficient unrestricted template survived from normal DNA to produce large quantities of product after 35 cycles of amplification.

After TaqI digestion of PCR products, in the standard PCR, the mutant heterozygote and normal controls showed the expected band pattern (Fig. 1A). An uncut band signifying the presence of a mutant sequence is visible in dilutions of E80K from normal up to 1:10. In the mutant enrichment PCR, the mutant heterozygote controls shows a very strong uncut band in the duplicate PCRs, with very faint cut bands resulting from residual unrestricted normal-strand templates remaining after the TaqI digestion in the early cycles of the PCR (Fig. 1B). The normal controls show faint products. The cut bands result from faithful amplification of normal template remaining after TaqI digestion, and the very faint uncut band is the result of polymerase infidelity on residual unrestricted normal templates. A very strong uncut band identifies the presence of E80K mutant up to a 1:100 dilution of mutantnormal DNA.

A set of 96 DNA samples of familial hypercholesterolemia patients from the Manchester area were used in a demonstration of the efficacy of the forced TaqI PCR/mutant enrichment protocol in identification of E80K heterozygous mutants in pooled samples. By using as PCR templates 8 samples each pooled from 12 columns and 12 samples each pooled from 8 rows of DNA samples contained in a standard 96-well array, it was possible to identify the coordinates of all possible E80K heterozygous mutants with 20 PCRs instead of 96 individual reactions, leaving a small number of possibilities to test by standard forced PCR and digest and confirm by sequencing (results not shown).

In conclusion, screening for CpG mutations in pooled DNA samples is possible by using CpG-PCR. We have detected a heterozygous E80K mutant in 100-fold excess of normal DNA after selective enrichment; i.e., the mutant allele represents ~0.5% of the total amplified sequences. Because the method is capable of identifying any single-base mutations within the four-base TaqI recognition site TCGA, the precise base change needs to be confirmed by direct sequencing. The approach is particularly suited to high-volume genotyping or screening programs by using
the MADGE system [8–10]. Column and row pooling from 96-well (8 x 12) sample arrays was used here, although 10- to 100-fold pooling would have been possible for detection of E80K. In general, n-dimensional pooling of pool size m reduces m^n single analyses to nm analyses. Efficiency for unique identification of mutants in arrays from pool coordinates alone is theoretically achieved by choosing n and m on the basis of expected mutant frequency and empirically possible pooling ratio [11, 12]. For example, three-dimensional 30-fold pooling of a 30 x 30 x 30 “array” would reduce screening of 27 000 samples for a rare CpG variant from 27 000 PCRs and restriction digests from 300 96-well plates of pools to one 96-well PCR plate and a single MADGE gel analysis. This procedure will represent an important savings in time and reagents in population studies where mutations at a CpG site of interest are expected at low frequency.

S.D.O’D. is funded by Medical Research Council Project Grant G9516890MB awarded to I.N.M.D. and S.E.H. D.J.W. is funded by the British Heart Foundation Group Grant RG/93008 awarded to S.E.H. S.E.H. funded by British Heart Foundation Chair and Program Grant RG/93008 awarded to I.N.M.D. and S.E.H. S.E.H. is funded by the British Heart Foundation Research Grant G9516890MB awarded to I.N.M.D. and S.E.H. S.D.O’D. is funded by Medical Research Council Project Grant PG95007. I.N.M.D. is a Lister Institute Professor.

References